We agree with Dr. Kuller (1) that large-scale epidemiologic studies require substantial input from innovative, committed investigators across a range of disciplines—as in the UK Biobank—in deciding what questions to ask, what measurements to make, and what samples to collect, so that a wide range of hypotheses about many different conditions may be addressed (2). Given the scientific potential of the UK Biobank and the controversies surrounding its initiation (3), it is hard to imagine a stronger vested interest in its success for the investigators involved. We also agree that detailed phenotyping in large prospective cohort studies can allow more specific questions to be addressed (4).

We disagree, however, that large-scale studies are easy to conduct or that data collection in the measures used is necessarily simple. For example, in the UK Biobank, touch-screen systems were developed to increase the precision and breadth of information that could be collected on such a large scale. Food frequency questions have been supplemented by Web-based diet diary information, and physical activity questions have been supplemented by wrist-worn activity monitors. Through careful design and in the context of specific hypotheses, retinal imaging has been obtained for 100,000 UK Biobank participants, and magnetic resonance imaging, dual-energy X-ray absorptiometry, and 3-dimensional carotid ultrasonography are being planned for 100,000 UK Biobank participants, and magnetic resonance imaging, dual-energy X-ray absorptiometry, and 3-dimensional carotid ultrasonography are being planned for 100,000 of them. All of these measures are designed by dedicated groups of actively involved experts, and data are collected using the same centralized approach to which Dr. Kuller objects. No one is paid on the basis of how many pieces of data they submit or per paper published. If there is a “top-down” (centralized?) approach, it is limited to the mechanics of establishing and running assessment centers and ensuring high-quality data collection. Few imaginative scientists have either an interest or expertise in the large-scale industrial processes needed to make such efforts successful; indeed, the most effective cohort investigators are often those who find capable managers to conduct such procedures for them. The successful merging of academic and management disciplines has been a major, and possibly unique, factor in the success of the UK Biobank.

Common diseases may be caused by many exposures that interact with each other in complex ways. To study the effects of numerous different exposures, large studies involving many thousands of cases of a specific disease are likely to be required (5). One area of phenotyping that does not typically receive sufficient focus is ascertainment of a wide range of health outcomes that occur during follow-up. Misclassification of disease cases can substantially reduce statistical power to assess associations. Hence, in the UK Biobank, as in many other cohort studies, diagnostic accuracy is increased by using innovative methods (e.g., retrieval of imaging data and tissue samples) for the classification of outcomes.

Establishing such epidemiologic resources requires more—not less—than would a study with a narrow focus on a few specific hypotheses. Prospective studies typically collect either large amounts of data on relatively few participants (“data depth”) or small amounts of data on many participants (“data breadth”). The centralized model we espoused (6), exemplified by Sweden’s LifeGene, the UK Biobank, and other efforts, can achieve both data depth and data breadth in an efficient manner. In contrast with Dr. Kuller’s assertion, “streamlined” does not mean simple, and “centralized” does not mean excluding scientists from the process. Rather, it means allowing scientists to focus on being innovative instead of distracting them with process.

The most serious challenge to epidemiology today is the need for very large cohort studies which produce high-quality data that are widely available to the scientific community, for a broad range of uses, at an affordable cost. The centralized model is one proven way of achieving this, to help identify the “big winners” we all seek.
ACKNOWLEDGMENTS

Author affiliations: National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland (Teri A. Manolio); and Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, Oxford, United Kingdom (Rory Collins).
Conflict of interest: none declared.

REFERENCES